Genomics Data 7 (2016) 318–320

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Deep RNA-Seq analysis reveals unexpected features of human prostate basal epithelial cells

Dingxiao Zhang [a,*], Daechan Park [b], Yue Lu [a], Jianjun Shen [a], Vishwanath R. Iyer [b], Dean G. Tang [a,c,*]

[a] Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Smithville, TX 78957, United States
[b] Institute for Cellular and Molecular Biology, Center for Systems and Synthetic Biology, and Section of Molecular Genetics and Microbiology, University of Texas at Austin, Austin, TX 78712-0159, United States
[c] Centers for Cancer Epigenetics, Stem Cell and Developmental Biology, RNA Interference and Non-Coding RNAs and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Houston, TX, United States

ARTICLE INFO

Article history:
Received 30 August 2015
Accepted 6 September 2015
Available online 10 September 2015

Keywords:
Prostate epithelial cells
Basal cells
Luminal cells
RNA-seq

ABSTRACT

Prostate cancer is the second leading cause of cancer-related deaths among American men [1]. The prostate gland mainly contains basal and luminal cells, which are constructed as a pseudostratified epithelium. Annotation of prostate epithelial transcriptomes provides a foundation for discoveries that can impact disease understanding and treatment. Here, for the first time, we describe a whole-genome transcriptome analysis of human benign prostatic basal and luminal populations by using deep RNA sequencing (GSE67070) [2]. Combined with comprehensive molecular and biological characterizations, we show that the differential gene expression profiles account for their distinct functional phenotypes. Strikingly, in contrast to luminal cells, basal cells preferentially express gene categories associated with stem cells, neural and neuronal development, and RNA processing. Of clinical relevance, the treatment failed castration-resistant and anaplastic prostate cancers molecularly resemble a basal-like phenotype. We also identified genes associated with patient clinical outcome. Therefore, we provide a gene expression resource for understanding human prostate epithelial lineages, and link the cell-type specific gene signatures to subtypes of prostate cancer development.

Specifications	
Organism/cell line/tissue	Homo sapiens/human prostate
Sex	Male
Sequencer or array type	Illumina HiSeq 2000
Data format	Raw and processed
Experimental factors	Freshly purified basal and luminal epithelial cells from benign human prostate
Experimental features	RNA-seq analysis of FACS-sorted human prostate basal and luminal cells derived from benign prostate tissues from patients with prostate cancer.
Consent	N/A
Sample source location	St. David's North Austin Medical Center, Austin, TX, USA

1. Direct link to deposited data

Deep RNA-seq: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE67070.

2. Experimental design, materials and methods

2.1. Human primary prostate tissue processing and FACS

All primary human prostate benign tissue samples were obtained with the written informed consent form from the patients with prostate cancer in accordance with federal and institutional guidelines and with the approved IRB protocols (MDACC LAB04-0498). The tissue processing protocol was previously described [3]. Briefly, collagenase (type I) and dispase were used to digest the prostate tissues followed by trypsinization to release single cells. The final dissociated single cell suspension was stained with PE-CD49f, APC-Cy7-CD45 (eBioscience, San Diego, CA) and APC-Trop2 antibodies (R&D Systems, Minneapolis, MN). Fluorescence-activated cell sorting (FACS) analysis and sorting were performed by using the BD Aria or Fusion (BD Biosciences, San Jose, CA). Propidium iodide was added prior to FACS analysis to separate

* Corresponding authors.
E-mail addresses: dzhang5@mdanderson.org (D. Zhang), dtang@mdanderson.org (D.G. Tang).

http://dx.doi.org/10.1016/j.gdata.2015.09.003

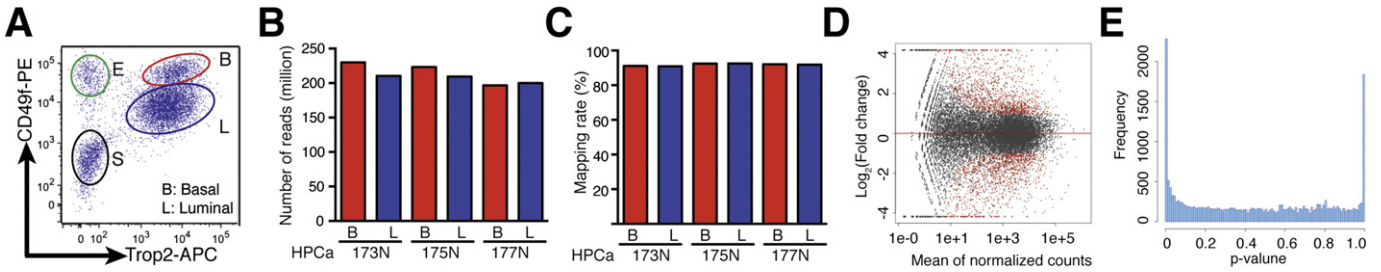

Fig. 1. A. FACS plots of human prostate cell lineage fractionation. Prostate basal (B), luminal (L), endothelial-enriched (E) and stromal-enriched (S) populations are identified as Trop2^{+-} CD49f^{hi}, Trop2^{+}CD49f^{lo}, Trop2^{-}cd49f^{hi} and Trop2^{-}CD49f^{-}, respectively. B. The number of sequencing reads generated for each sample. C. High mapping rate indicates the high quality of our RNA-seq data. D. MA plot showing the appropriate normalization of RNA-seq data. The red dots represent the genes with p-value < 0.05. E. Histogram showing the distribution of p-values.

viable from dead cells. In this study, three representative human benign samples (HPCa173N, 175 N, 177 N) from PCa patients who had largely benign biopsies were selected. After excluding the CD45^{+} immune cells, prostate epithelial cells were isolated as Trop2^{+}CD49f^{hi} (basal-enriched population) and Trop2^{+}CD49f^{lo} (luminal-enriched population) fractions (Fig. 1A).

2.2. RNA isolation

The FACS purified human prostate basal (Trop2^{+}CD49f^{hi}) and luminal (Trop2^{+}CD49f^{lo}) epithelial populations were subject to total RNA extraction by RNeasy mini kit (Qiagen, Valencia, CA) according to the manufacturer's instruction. Of note, the DNase digestion step was included to eliminate genomic DNA contamination. RNA extracts were assessed for quality by an Agilent 2100 Bioanalyzer, samples with A260/280 (2.0 ± 0.2) and RNA integrity number (RIN) ≥9.2 were used for further RNA-seq analysis.

2.3. Deep RNA-seq analysis and bioinformatics

cDNA libraries were constructed by using a TruSeq Stranded Total RNA Preparation Kit (Illumina, cat#: RS-122-2301). Of note, this kit contains the component of Ribo-ZeroTM Gold, which allows the depletion of ribosome RNA. Importantly, we only amplified our libraries with 10 PCR cycles instead of 15 cycles suggested by the manufacturer's manual to minimize duplicates that arise from PCR amplification. Purified libraries were further quantified using a Kapa library quantification kit (KAPA Biosystems, Boston, MA), and then loaded on cBot (Illumina, San Diego, CA) at final concentration of 10 pM to perform cluster generation, followed by 2 × 76 bp sequencing on HiSeq 2000 (Illumina, San Diego, CA).

In this study, every two libraries were pooled and loaded in one lane of HiSeq 2000, producing an average of 400 million 75-mer reads per lane. From each sample, we obtained an average of 211.5 M reads per sample (range from 196.6 to 229.9 M), indicating the high depth of sequencing (Fig. 1B). We mapped the sequencing reads to the reference human genome sequence (NCBI 36.1 [hg19] assembly) using TopHat v2.0.9 and Bowtie v2.1.0. Due to the use of rRNA-depleted total RNA for library construction, we expected high mapping rate of the reads. Indeed, we observed an average mapping rate of 91.7% (range from 90.8% to 92.4%) (Fig. 1C). Then, we assembled the alignments into gene transcripts and calculated their relative abundance using Cufflinks v2.1.1 and HTSeq v0.5.3p9. DESeq v1.10.1 was employed as a statistical procedure to call differentially expressed genes (DEGs) in different samples. For Gene Ontology (GO) analysis, IPA (Qiagen, Valencia, CA) and DAVID version 6.7 were used with human gene symbols. GSEA was carried out by using the curated gene sets (C2) of the Molecular Signature Database (MSigDB) version 4.0 provided by the Broad Institute (http://www.broad.mit.edu/gsea/).

Our deep RNA sequencing allows the detection of expression of more then 19,813 genes (normalized read counts >20). The MA plot and histogram of p-values distribution reveal the high quality of our RNA-seq data (Fig. 1D and E). Moreover, unsupervised hierarchical clustering showed that the basal and luminal populations were grouped together and well separated [2], demonstrating the biological reproducibility. To further validate whether these transcriptomic profiles were representative, we plotted the expression of known phenotypic markers specific to each lineage. As expected, prostatic basal and

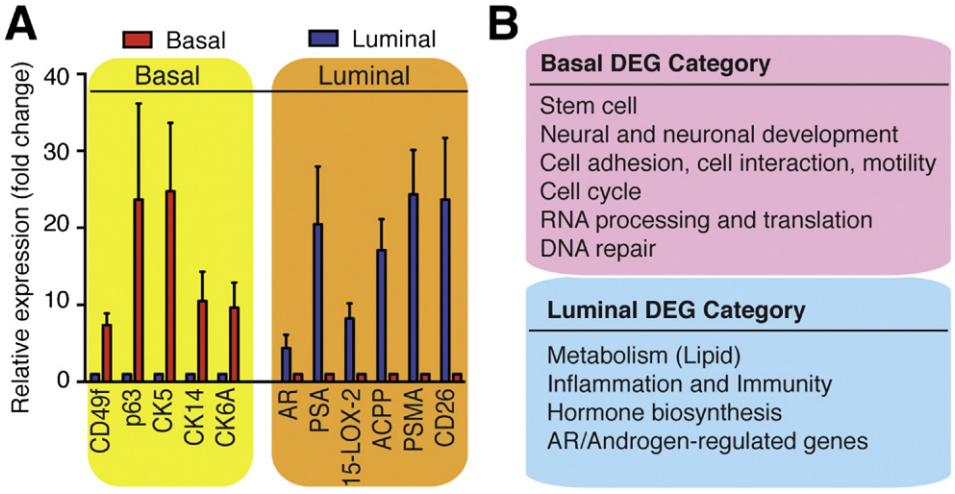

Fig. 2. A. Bar graph presentation of expression of known phenotypic markers corresponding to different prostate epithelial lineages. Gene expression values were extracted from RNA-seq data and expressed as relative fold changes. B. Distinct transcriptomic profiles of human prostatic basal and luminal cells. Shown are the gene categories overrepresented in basal (compared to luminal, up) and luminal (compared to basal, down) cell populations.

luminal markers showed exclusively high expression in their corresponding population (Fig. 2A). Together, this established the repeatability and representativeness of our RNA-Seq data.

Based on the systematic analysis of basal and luminal DEGs by using a combination of bioinformatics tools (e.g. GO, IPA, GSEA), we uncovered unexpected features for each epithelial cell lineage (Fig. 2B). Basal and luminal DEGs by DESeq with adjusted P-value (i.e., FDR) cut-off 0.05 were queried in DAVID and IPA for pathway analysis. Additionally, GSEA was performed with the genes ranked by signal-to-noise ratio of basal to luminal. These functional analyses uncovered novel features for each epithelial cell lineage. Basal cells preferentially expressed genes associated with categories of stem cells, cell adhesion and motility, neural and neuronal development, and RNA metabolism and ribosome/protein translation. In contrast, genes upregulated in luminal cells fell into categories of prostate organ function (e.g., steroid hormone processing, secretion), androgen and androgen receptor signaling, inflammation and immunity, and lipid metabolism. According to their distinct gene expression profiles, and consistent with the long-thought biological roles that the two different lineages play, our data indicate that the basal cell population harbors undifferentiated stem-like cells, and luminal layer represents the differentiated functional and secretory cells, respectively. The interpretation and description of these data can be found in our recent publication [2].

2.4. Validation of RNA-Seq data by qRT-PCR and immunostaining

Our RNA-Seq data was validated at both mRNA and protein levels for many representative genes. The first-strand cDNA synthesis was achieved by reverse transcription of RNA using random hexamers and SuperScript III Reverse Transcriptase (Invitrogen, Carlsbad, CA). Quantitative RT-PCR was performed using the iQ™ SYBR® Green supermix (BioRad, Hercules, CA) on a 7900HT Fast Real-Time PCR System (ABI, Applied Biosystems, Foster City, CA). Several phenotypic marker genes (e.g. CK5 and CK14 for basal cells, AR and PSA for luminal cells) were confirmed by qRT-PCR. For genes of interest, immunohistochemistry (IHC) and/or immunofluorescence (IF) staining were utilized to demonstrate their cellular localizations in human prostate tissues [2]. All staining were performed on either 5-μm paraffin or OCT frozen sections. Basic IF procedures have been described previously [4,5]. Images were captured by Olympus IX71 and Zeiss LSM510 META confocal microscopes.

3. Discussion

Biologically, gene expression is a key determinant of cellular phenotypes. A comprehensive annotation of transcriptome facilitates a better understanding of how gene expression influences phenotypic manifestations. Herein we describe the first whole-genome transcriptional profiling of human prostate basal and luminal epithelial cells by deep RNA-Seq. Through comprehensive molecular and biological characterizations, we have uncovered several unexpected features for human prostate basal and luminal cells [2]. Being currently the largest-scale gene expression profiling for human prostatic epithelial lineages, this dataset should provide valuable insights into understanding of prostate epithelial biology, as well as prostate cancer etiology [1].

Disclosures

All authors claim no conflicts of interests.

References

[1] American Cancer Society, Global Cancer Facts & Figures. American Cancer Society, Atlanta, 2015.
[2] D. Zhang, et al., Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. Nat. Commun. 7 (2016) 10798, http://dx.doi.org/10.1038/ncomms10798.
[3] A.S. Goldstein, et al., Purification and direct transformation of epithelial progenitor cells from primary human prostate. Nat. Protoc. 6 (2011) 656–667.
[4] C.R. Jeter, et al., Functional evidence that the self-renewal gene NANOG regulates human tumor development. Stem Cells 27 (2009) 993–1005.
[5] D. Zhang, et al., Arginine and glutamate-rich 1 (ARGLU1) interacts with mediator subunit 1 (MED1) and is required for estrogen receptor-mediated gene transcription and breast cancer cell growth. J. Biol. Chem. 286 (2011) 17746–17754.